

Peeling potatoes near-optimally in near-linear time

Sergio Cabello* Josef Cibulka† Jan Kynčl‡ Maria Saumell§

Pavel Valtr¶

June 6, 2014

Abstract

We consider the following geometric optimization problem: find a convex polygon of maximum area contained in a given simple polygon P with n vertices. We give a randomized near-linear-time $(1 - \varepsilon)$ -approximation algorithm for this problem: in $O((n/\varepsilon^4) \log^2 n \log(1/\delta))$ time we find a convex polygon contained in P that, with probability at least $1 - \delta$, has area at least $(1 - \varepsilon)$ times the area of an optimal solution.

Keywords: geometric optimization; potato peeling; visibility graph; geometric probability; approximation algorithm.

*Department of Mathematics, IMFM, and Department of Mathematics, FMF, University of Ljubljana, Slovenia. Supported by the Slovenian Research Agency, program P1-0297, projects J1-4106 and L7-5459, and by the ESF EuroGIGA project (project GReGAS) of the European Science Foundation.

†Department of Applied Mathematics and Institute for Theoretical Computer Science, Charles University, Faculty of Mathematics and Physics, Czech Republic. Supported by the project CE-ITI (GAČR P202/12/G061) of the Czech Science Foundation.

‡Department of Applied Mathematics and Institute for Theoretical Computer Science, Charles University, Faculty of Mathematics and Physics, Czech Republic; and Alfréd Rényi Institute of Mathematics, Hungary. Supported by the project CE-ITI (GAČR P202/12/G061) of the Czech Science Foundation and by ERC Advanced Research Grant no 267165 (DISCONV).

§Department of Mathematics, University of West Bohemia, Czech Republic. Supported by projects NEXLIZ - CZ.1.07/2.3.00/30.0038, which is co-financed by the European Social Fund and the state budget of the Czech Republic, and ESF EuroGIGA project ComPoSe as F.R.S.-FNRS - EUROGIGA NR 13604.

¶Department of Applied Mathematics and Institute for Theoretical Computer Science, Charles University, Faculty of Mathematics and Physics, Czech Republic. Supported by the project CE-ITI (GAČR P202/12/G061) of the Czech Science Foundation.

1 Introduction

We consider the algorithmic problem of finding a maximum-area convex set in a given simple polygon. Thus, we are interested in computing

$$A^*(P) := \sup\{\text{area}(K) \mid K \subset P, K \text{ convex}\}.$$

The problem was introduced by Goodman [22], who named it the *potato peeling problem*. Goodman also showed that the supremum is actually achieved, so we can replace it by the maximum. Henceforth we use n to denote the number of vertices in the input polygon P .

Chang and Yap [10] showed that $A^*(P)$ can be computed in $O(n^7)$ time. Since there have been no improvements in the running time of exact algorithms, it is natural to turn the attention to faster, approximation algorithms. A step in this direction is made by Hall-Holt et al. [24], who show how to obtain a constant-factor approximation in $O(n \log n)$ time.

In this paper we present a randomized $(1 - \varepsilon)$ -approximation algorithm. Besides the simple polygon P , the algorithm takes as input a parameter $\varepsilon \in (0, 1)$ controlling the approximation and a parameter $\delta \in (0, 1)$ controlling the probability of failure. In time $O((n/\varepsilon^6) \log^2 n \log(1/\delta))$ the algorithm returns a convex polygon contained in P that, with probability $1 - \delta$, has area at least $(1 - \varepsilon) \cdot A^*(P)$. For any constant ε and δ , the running time becomes $O(n \log^2 n)$.

Overview of the approach. Let R be a set of points contained in P . The *visibility graph* of R , denoted by $G(P, R)$, has R as vertex set and, for any two points x and y in R , the edge xy is in $G(P, R)$ whenever the segment xy is contained in P . See Figure 1.

Let us assume that the set of points R is obtained by uniform sampling in P . We note the following properties:

- For each convex polygon $K \subseteq P$, the area of the convex hull $CH(K \cap R)$ is similar to the area of K , provided that $|K \cap R|$ is large enough. For this, it is convenient to have large $|R|$.
- For each convex polygon $K \subseteq P$, the boundary of $CH(K \cap R)$ is made of edges in $G(P, R)$.
- With dynamic programming one can find a maximum-area convex polygon defined by edges of $G(P, R)$. For this to be efficient, it is convenient that $G(P, R)$ has few edges.

Thus, we have a trade-off on the number of points in R that are needed. We argue that there is a suitable size for R such that $G(P, R)$ has a near-linear expected number of edges and, with reasonable probability, the edges of $G(P, R)$ give a good inner approximation to an optimal solution. Instead of finding the optimal solution directly in $G(P, R)$, we make a search in a small parallelogram of area $\Theta(A^*(P))$ around each edge of $G(P, R)$, performing a second sampling. The core of the argument is a bound relating $A^*(P)$ and the probability that two random points in P are visible. Such relation was unknown and we believe that it is of independent interest.

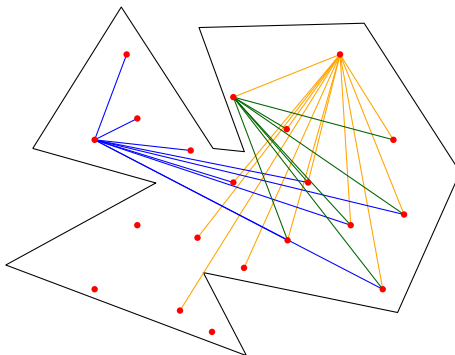


Figure 1: A portion of the visibility graph of a point set. Only the edges incident to three vertices are displayed.

Other related work. There have been several results about finding maximum-area objects of certain type inside a given simple polygon. DePano, Ke and O’Rourke [17] consider squares and equilateral triangles, Daniels, Milenkovic and Roth [16] consider axis-parallel rectangles, Melissaratos and Souvaine [26] consider arbitrary triangles. Sub-quadratic algorithms to find a longest segment contained in a simple polygon were first given by Chazelle and Sharir [14] and improved by Agarwal, Sharir and Toledo [1, 2].

Aronov et al. [3] consider a variation where the search is restricted to convex polygons whose edges are edges of a given triangulation (with inner points) of P . They show how to compute a maximum-area convex polygon for this model in $O(m^2)$ time, where m is the number of edges in the triangulation.

Dumitrescu, Har-Peled and Tóth [18] consider the following problem: given a unit square Q and a set X of points inside Q , find a maximum-area convex body inside Q that does not have any point of X in its interior. This is an instance of the potato peeling problem for polygons *with holes*. In the full version, they provide a $(1 - \varepsilon)$ -approximation in time $O((n^2/\varepsilon^8) \log^2 n)$; the conference version has slightly worse time bounds. For any fixed ε , the running time is near-quadratic. Our algorithm exploits the absence of holes in P , so it does not produce an improvement in this case.

The potato peeling problem can be understood as finding a largest set of points that are mutually visible. Rote [29] showed how to compute in polynomial time the probability that two random points inside a polygon are visible. Cheong, Efrat and Har-Peled [15] consider the problem of finding a point in a simple polygon whose visibility region is maximized. They provide a $(1 - \varepsilon)$ -approximation algorithm using near-quadratic time. The approach is based on taking a random sample of points in the polygon, constructing the visibility region of each point, and taking a point laying in most visibility regions.

Roadmap. In Section 2 we provide tools related to convex bodies. In Section 3 we relate the probability of two random points being visible and $A^*(P)$. Finally, we present and analyze the algorithm in Section 4. We conclude in Section 5.

Assumptions. We assume that $\varepsilon > 1/n$. Otherwise we can solve the problem in time $O(n^7) = O(n/\varepsilon^6)$ using the algorithm of Chang and Yap [10]. We will have to generate points uniformly at random inside a triangle. For this, we will assume that a random number in the interval $[0, 1]$ can be generated in constant time.

2 About convexity

Here we provide tools related to convexity. In the first part, we provide results about the number of points that have to be sampled inside a convex body K so that the area of the convex hull of the sample is a good approximation to the area of K . We also provide extensions to the case where we sample points in a superset of K . In the second part we give an algorithm to find a largest convex polygon whose edges are defined by a visibility graph inside a polygon.

2.1 Inner approximation using random sampling

Lemma 1. *Let K be a convex body in the plane and let R be a sample of points chosen uniformly at random inside K . There is some universal constant C_1 such that, if $|R| \geq C_1/\varepsilon^{3/2}$, then with probability at least $5/6$ it holds that $\text{area}(CH(R)) \geq (1 - \varepsilon) \cdot \text{area}(K)$.*

Proof. We use as a black box known extremal properties and bounds on the so-called missed area of a random polygon. See the lectures by Bárány [4, 2nd lecture], the survey [5] or [6] for an overview.

Let us scale K such that it has area 1. We have to show that $1 - \text{area}(CH(R)) \geq \varepsilon$ holds with probability at most $1/6$.

Let K_m denote the convex hull of m points chosen uniformly at random in K and define $X(m) = 1 - \text{area}(K_m)$. Thus $X(m)$ is the *missed area*, that is, the area of $K \setminus K_m$. Groemer [23] showed that $\mathbb{E}[X(m)]$ is maximized when K is a disk of area 1. Rényi and Sulanke [28] showed that for every smooth convex set K there exists some constant C_K , depending on K , such that $\mathbb{E}[X(m)] = C_K \cdot m^{-2/3}$. This result also follows from a similar upper bound by Rényi and Sulanke [27] on the expected number E_m of edges of K_m and from Efron's [20] identity $\mathbb{E}[X(m)] = \mathbb{E}[E_{m+1}]/(m+1)$. Both statements together imply that

$$\mathbb{E}[X(m)] \leq \frac{C'}{m^{2/3}},$$

where C' is the constant C_K when K is a unit-area disk. (From the results of [28], or subsequent works, one can explicitly compute that $C' \leq 5$, so the constant is very reasonable.)

We set $C_1 = (6C')^{3/2}$. Whenever $|R| \geq C_1 \cdot \varepsilon^{-3/2}$, we can use Markov's inequality to obtain

$$\begin{aligned} \Pr[1 - \text{area}(CH(R)) \geq \varepsilon] &= \Pr[X(|R|) \geq \varepsilon] \\ &\leq \frac{\mathbb{E}[X(|R|)]}{\varepsilon} \\ &\leq \frac{C' |R|^{-2/3}}{\varepsilon} \\ &\leq \frac{C' ((6C')^{3/2} \cdot \varepsilon^{-3/2})^{-2/3}}{\varepsilon} \\ &= \frac{1}{6}. \end{aligned}$$

□

Note that for any $\varepsilon \in (0, 1)$ it must be that $C_1/\varepsilon^{3/2} \geq 3$, as otherwise $\text{area}(CH(R)) = 0$.

Lemma 2. *Let K be a convex body contained in a polygon P , let R be a random sample of points inside P , and let $C \geq 3$ be an arbitrary value. If*

$$|R| \geq 4 \cdot C \cdot \frac{\text{area}(P)}{\text{area}(K)},$$

then with probability at least $5/6$ it holds that $|R \cap K| \geq C$.

Proof. Let $X = |R \cap K|$. The random variable X is a sum of $|R|$ Bernoulli independent random variables, where each Bernoulli random variable has expected value

$$p = \frac{\text{area}(K)}{\text{area}(P)}.$$

Standard calculations (or formulas) show that

$$\mathbb{E}[X] = |R| \cdot p \geq 4 \cdot C \cdot \frac{\text{area}(P)}{\text{area}(K)} \cdot \frac{\text{area}(K)}{\text{area}(P)} = 4 \cdot C$$

and

$$\text{Var}[X] = |R| \cdot p(1-p) \leq \mathbb{E}[X].$$

We can now use Chebyshev's inequality in its form

$$\forall a > 0 : \Pr[|X - \mathbb{E}[X]| \geq a] \leq \frac{\text{Var}[X]}{a^2}$$

and the inequality $C \geq 3$ to obtain the following

$$\begin{aligned} \Pr[X \leq C] &\leq \Pr[X \leq \tfrac{1}{4}\mathbb{E}[X]] \\ &\leq \Pr[|X - \mathbb{E}[X]| \geq \tfrac{3}{4}\mathbb{E}[X]] \\ &\leq \frac{4^2}{3^2} \cdot \frac{\text{Var}[X]}{(\mathbb{E}[X])^2} \\ &\leq \frac{16}{9} \cdot \frac{1}{\mathbb{E}[X]} \\ &\leq \frac{16}{9} \cdot \frac{1}{4 \cdot C} \\ &\leq \frac{16}{9} \cdot \frac{1}{4 \cdot 3} \\ &< \frac{1}{6}. \end{aligned}$$

□

Lemma 3. *Let K be a convex body contained in a polygon P , let R be a random sample of points inside P , and let C_1 be the constant in Lemma 1. If*

$$|R| \geq 4 \cdot \frac{C_1}{\varepsilon^{3/2}} \cdot \frac{\text{area}(P)}{\text{area}(K)},$$

then with probability at least $2/3$ it holds that $\text{area}(CH(R \cap K)) \geq (1 - \varepsilon) \text{area}(K)$.

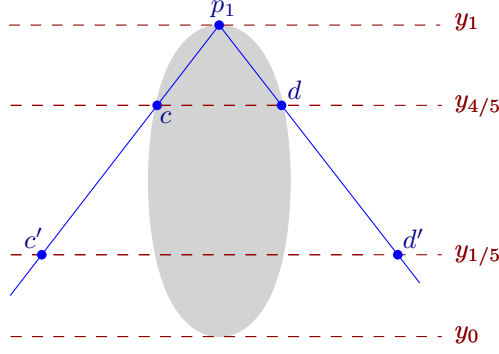


Figure 2: Proof of Lemma 4.

Proof. We define the following events:

$$\begin{aligned}\mathcal{E} : \quad |R \cap K| &\geq C_1/\varepsilon^{3/2}, \\ \mathcal{F} : \quad \text{area}(CH(R \cap K)) &\geq (1 - \varepsilon) \cdot \text{area}(K).\end{aligned}$$

For each event \mathcal{A} we use $\overline{\mathcal{A}}$ for its negation. Since $C_1/\varepsilon^{3/2} \geq 3$ (see the remark after Lemma 1), Lemma 2 implies

$$\Pr[\mathcal{E}] \leq \frac{1}{6}.$$

and Lemma 1 implies

$$\Pr[\overline{\mathcal{F}} \mid \mathcal{E}] \leq \frac{1}{6}.$$

Therefore

$$\begin{aligned}\Pr[\overline{\mathcal{F}}] &= \Pr[\overline{\mathcal{F}} \mid \mathcal{E}] \cdot \Pr[\mathcal{E}] + \Pr[\overline{\mathcal{F}} \mid \overline{\mathcal{E}}] \cdot \Pr[\overline{\mathcal{E}}] \\ &\leq \Pr[\overline{\mathcal{F}} \mid \mathcal{E}] + \Pr[\overline{\mathcal{E}}] \\ &\leq \frac{1}{6} + \frac{1}{6} \\ &= \frac{1}{3}.\end{aligned}$$

□

2.2 Outer containment in a parallelogram

Let K be a convex body in \mathbb{R}^2 . We use $y(p)$ to denote the y -coordinate of a point p . For each $\alpha \in (0, 1)$ we define $y_\alpha(K)$ as the unique value satisfying

$$\text{area}(\{p \in K \mid y(p) \leq y_\alpha(K)\}) = \alpha \cdot \text{area}(K).$$

Thus, the horizontal line at height $y_\alpha(K)$ breaks K into two polygons and the lower one has a proportion α of the area of K . Let us further define

$$y_0 = \min\{y(p) \mid p \in K\} \quad \text{and} \quad y_1 = \max\{y(p) \mid p \in K\}.$$

Lemma 4. *For each convex body K*

$$y_1(K) - y_{4/5}(K) \leq y_{4/5}(K) - y_{1/5}(K) \quad \text{and} \quad y_{1/5}(K) - y_0(K) \leq y_{4/5}(K) - y_{1/5}(K).$$

Proof. In this proof, let us drop the dependency on K in the notation and set $y_\alpha = y_\alpha(K)$ for each $\alpha \in [0, 1]$. Without loss of generality we assume that $\text{area}(K) = 1$. We only show that $y_1 - y_{4/5} \leq y_{4/5} - y_{1/5}$; the other inequality is symmetric.

Let p_1 be a highest point of K , let cd be the intersection of the line $y = y_{4/5}$ with K , let ℓ_c be the line through p_1 and c , let ℓ_d be the line through p_1 and d , let c' be the intersection of ℓ_c with $y = y_{1/5}$, and let d' be the intersection of ℓ_d with $y = y_{1/5}$. See Figure 2. Since the triangle p_1cd is contained in K , its area is smallest than $1/5$ and thus

$$\text{area}(p_1cd) = \frac{1}{2} \cdot |cd| \cdot (y_1 - y_{4/5}) \leq \frac{1}{5}.$$

Because K is convex, the portion of K between $y = y_{4/5}$ and $y = y_{1/5}$ lies in the trapezoid $aa'b'b$, which must have area at least $3/5$. Therefore

$$\frac{3}{5} \leq \text{area}(aa'b'b) = \frac{1}{2} \cdot (|cd| + |c'd'|) \cdot (y_{4/5} - y_{1/5}).$$

Using that

$$\frac{c'd'}{cd} = \frac{y_1 - y_{1/5}}{y_1 - y_{4/5}} = \frac{y_1 - y_{4/5} + y_{4/5} - y_{1/5}}{y_1 - y_{4/5}} = 1 + \frac{y_{4/5} - y_{1/5}}{y_1 - y_{4/5}}$$

we have

$$\frac{3}{5} \leq \text{area}(aa'b'b) = \frac{1}{2} \cdot |cd| \cdot \left(2 + \frac{y_{4/5} - y_{1/5}}{y_1 - y_{4/5}}\right) \cdot (y_{4/5} - y_{1/5}).$$

If we would have $y_1 - y_{4/5} > y_{4/5} - y_{1/5}$ then the last relation would imply that

$$\begin{aligned} \frac{3}{5} &\leq \frac{1}{2} \cdot |cd| \cdot \left(2 + \frac{y_{4/5} - y_{1/5}}{y_1 - y_{4/5}}\right) \cdot (y_{4/5} - y_{1/5}) \\ &< \frac{1}{2} \cdot |cd| \cdot (2 + 1) \cdot (y_1 - y_{4/5}) \\ &= 3 \cdot \text{area}(p_1cd) \\ &= \frac{3}{5}. \end{aligned}$$

Therefore it must be that $y_1 - y_{4/5} \leq y_{4/5} - y_{1/5}$. □

For any two points a and b and any value $A \geq 0$, let $\Gamma(a, b, A)$ denote the parallelogram containing any horizontal translation of the segment $a'b'$ by distance at most $\frac{2A}{|y(a) - y(b)|}$, where $a' = 2a - b$ and $b' = 2b - a$. See Figure 3, left. Note that $|a'b'| = 2|ab|$ and $\text{area}(\Gamma(a, b, A)) = 12 \cdot A$.

Lemma 5. *Let K be a convex body and assume that $A \geq \text{area}(K)$. Let a and b be points in K such that*

$$y(a) \geq y_{4/5}(K) \quad \text{and} \quad y(b) \leq y_{1/5}(K)$$

Then K is contained $\Gamma(a, b, A)$.

Proof. In this proof, let us drop the dependency on K in the notation and set $y_\alpha = y_\alpha(K)$ for each $\alpha \in [0, 1]$.

Because of Lemma 4 we have

$$y(a') = 2y(a) - y(b) \geq y(a) + y_{4/5} - y_{1/5} \geq y(a) + y_1 - y_{4/5} \geq y_1$$

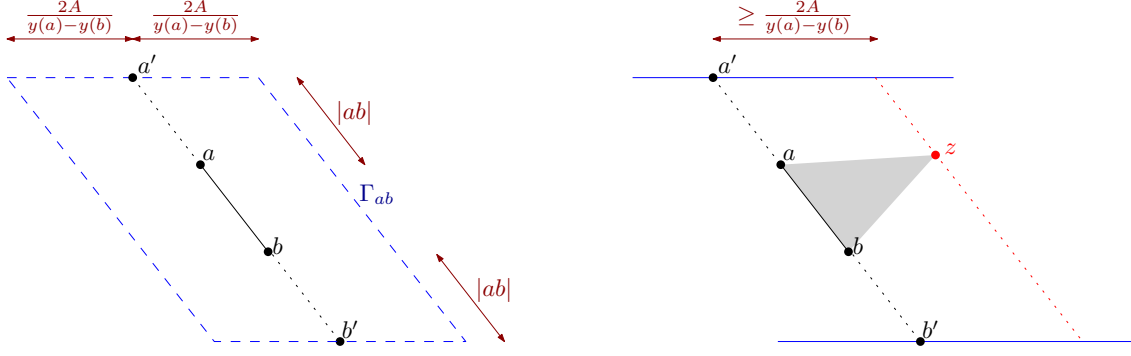


Figure 3: Left: parallelogram $\Gamma(a, b, A)$. Right: proof of Lemma 5.

and similarly

$$y(b') = 2y(b) - y(a) \leq y(b) + y_{1/5} - y_{4/5} \leq y(b) + y_0 - y_{1/5} \leq y_0.$$

Therefore K is contained between the horizontal lines $y = y(a')$ and $y = y(b')$. These are the lines supporting the top and bottom side of $\Gamma(a, b, A)$.

Assume, for the sake of a contradiction, that K has some point z outside Γ_{ab} . Since z lies between the lines $y = y(b')$ and $y = y(a')$, it must be that the horizontal distance from z to $a'b'$ is more than $\frac{2A}{y(a)-y(b)}$. See Figure 3, right. Since the triangle abz is contained in K we would have

$$\text{area}(K) \geq \text{area}(abz) > \frac{1}{2} \cdot (y(a) - y(b)) \cdot \frac{2A}{y(a) - y(b)} = A \geq \text{area}(K),$$

which is a contradiction. Therefore any point of K is contained in $\Gamma(a, b, A)$. \square

Lemma 6. *Let K be a convex body contained in a polygon P , and assume that $A \geq \text{area}(K)$. If R is a random sample of points inside P with*

$$|R| \geq 60 \cdot \frac{\text{area}(P)}{\text{area}(K)},$$

then with probability at least $2/3$ it holds that R contains two points a and b such that $\Gamma(a, b, A)$ contains K .

Proof. Define

$$K_{\leq 1/5} = \{p \in K \mid y(p) \leq y_{1/5}(K)\} \quad \text{and} \quad K_{\geq 4/5} = \{p \in K \mid y(p) \geq y_{4/5}(K)\},$$

and consider the following events:

$$\mathcal{E}_{\leq 1/5} : K_{\leq 1/5} \cap R \neq \emptyset \quad \text{and} \quad \mathcal{E}_{\geq 4/5} : K_{\geq 4/5} \cap R \neq \emptyset.$$

Since

$$|R| \geq 4 \cdot 3 \cdot \frac{\text{area}(P)}{\text{area}(K)/5} = 4 \cdot 3 \cdot \frac{\text{area}(P)}{\text{area}(K_{\leq 1/5})} = 4 \cdot 3 \cdot \frac{\text{area}(P)}{\text{area}(K_{\geq 4/5})},$$

Lemma 2 implies

$$\Pr[\mathcal{E}_{\leq 1/5}] \geq \frac{5}{6} \quad \text{and} \quad \Pr[\mathcal{E}_{\geq 4/5}] \geq \frac{5}{6}.$$

It follows that

$$\Pr[\mathcal{E}_{\leq 1/5} \cap \mathcal{E}_{\geq 4/5}] \geq \frac{2}{3}.$$

When $\mathcal{E}_{\leq 1/5}$ and $\mathcal{E}_{\geq 4/5}$ hold, there are points $a \in K_{\leq 1/5} \cap R$ and $b \in K_{\geq 4/5} \cap R$ and Lemma 5 implies that K is contained in $\Gamma(a, b, A)$. \square

2.3 Largest convex polygon in a visibility graph.

Let H be a visibility graph in some simple polygon. We denote the set of vertices and edges of H by $V(H)$ and $E(H)$, respectively. We assume that the coordinates of the vertices of H are known. A set of vertices U from H is a **convex clique** if: (i) there is an edge between any two vertices of U , and (ii) the points of U are in convex position. The **area of a convex clique** U is the area of $CH(U)$.

Let s be a point of $V(H)$. We are interested in finding a convex clique of maximum area in H , denoted by $\varphi(H, s)$, that has s as highest point. Thus we want

$$\varphi(H, s) \in \arg \max \{ \text{area}(U) \mid U \subseteq V(H) \text{ a convex clique, } s \text{ highest point in } U \}.$$

Lemma 7. *For any point s of $V(H)$, we can compute $\varphi(H, s)$ in time $O(|V(H)|^2)$.*

Proof. Pruning vertices, we can assume that all vertices of H are adjacent to s and below s . We can then use the algorithm of Bautista-Santiago et al. [7], which is an improvement over the algorithm of Fischer [21], restricted to the edges that are in H . For completeness, we provide a quick overview of the approach.

For this proof, let us denote $n = |V(H)| - 1$. We sort the points $V(H) \setminus \{s\}$ counterclockwise radially from s . Let x_1, x_2, \dots, x_n be the labeling of the points $V(H) \setminus \{s\}$ according to that ordering. Thus, for each $i < j$ the sequence x_i, s, x_j is a right turn.

Using a standard point-line duality and constructing the arrangement of lines dual to the points $V(H)$, we get the circular order of the edges around each point x_i [25]. For this we spend in total $O(n^2)$ time [13, 19].

For each $i < j$ such that $x_i x_j \in E(H)$, let $\text{OPT}[i, j]$ be the largest-area convex clique U that has x_i, x_j , and s *consecutively* along the boundary of $CH(U)$. We then have

$$\text{area}(\varphi(H, s)) = \max_{i < j, x_i x_j \in E(H)} \text{OPT}[i, j].$$

Taking the convention that $\max \emptyset = 0$, the values $\text{OPT}[i, j]$ satisfy the following recursion

$$\begin{aligned} \text{OPT}[i, j] &= \text{area}(ax_i x_j) \\ &+ \max \{ \text{OPT}[h, i] \mid h < i, x_h x_i \in E(H), x_h, x_i, x_j \text{ makes a left turn} \}. \end{aligned}$$

To argue the correctness of the recursion, one needs to observe that the right side of the equation does indeed correspond to the construction of a convex polygon.

For any fixed i , the values $\text{OPT}[i, *], * > i$, can be computed in $O(n)$ time, provided that the edges incident to x_i are already radially sorted and the values $\text{OPT}[h, i]$ are already available for all $h < i$. To achieve linear time, one performs a scan of the edges incident to x_i and uses the property that

$$\{x_h x_i \in E(H) \mid h < i, x_h, x_i, x_j \text{ makes a left turn}\}$$

forms a contiguous sequence in the circular ordering of edges incident to x_i . Thus, we can fill in the whole table $\text{OPT}[\cdot, \cdot]$ in time $O(n^2)$. With this we can compute $\text{area}(\varphi(H, s))$ and construct an optimal solution $\varphi(H, s)$ by standard backtracking. See [7] for additional details. \square

3 Probability for visibility

A polygon P is **weakly visible** from a segment s if, for each point $p \in P$, there exists some point $x \in s$ such that $xs \subset P$.

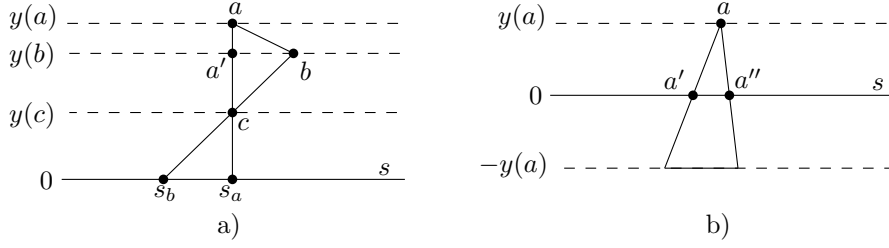


Figure 4: Situation in the proof of Theorem 8. a) Points a and b are on the same side of s . b) Points a and b are on different sides of s .

Theorem 8. *Let P be a unit-area polygon weakly visible from a diagonal s . Let a and b be two points chosen uniformly at random in P . Then*

- (i) $\Pr[ab \subset P] \leq 18 \cdot A^*(P)$ and
- (ii) $\Pr[ab \subset P \ \& \ ab \cap s \neq \emptyset] \leq 6 \cdot A^*(P)$.

Proof. Without loss of generality we assume that s is a horizontal segment on the x -axis. In this proof we use $y(a)$ to denote the y -coordinate of a point a . To simplify the notation, we use in this proof $A^* = A^*(P)$.

Consider first the point a fixed. We first bound the probability that a and b are visible and $|y(a)| \geq |y(b)|$ to obtain the following:

$$\Pr[ab \subset P \text{ and } |y(a)| \geq |y(b)|] \leq 9 \cdot A^*.$$

This is seen showing that the set of points b satisfying $ab \subset P$ and $|y(a)| \geq |y(b)|$ are within a region of area at most $9A^*$.

We distinguish two cases:

- 1) $y(a)y(b) \geq 0$ (a and b are on the same side of s).
- 2) $y(a)y(b) < 0$ (a and b are on the opposite sides of s).

Let us first consider case 1). Refer to Figure 4a). We know that a sees some point s_a on s . With a shear transformation that preserves the area, we can assume that the segment as_a is vertical. We know that b sees some point s_b on s . If b sees a , we can choose s_b so that as_a and bs_b share a common point (either the two segments intersect or $s_a = s_b$). Let c be the common point of as_a and bs_b . We have a generalized polygon $Q = abs_bs_a$ (two sides cross or one side has zero length) whose boundary is in P , and therefore the whole interior of Q is also in P . Here we use that P has no holes.

Let h be a horizontal line through b and let a' be the intersection between h and the segment as_a . The interior of Q is made of two triangles, abc and s_as_bc , both contained in P and thus each of them has area at most A^* .

For the triangle abc , we have $|ac| \cdot |a'b| \leq 2A^*$, which implies that

$$|a'b| \leq \frac{2A^*}{y(a) - y(c)}. \quad (1)$$

For the triangle s_as_bc , we have $y(c) \cdot |s_as_b| \leq 2A^*$. By the similarity of the triangles s_as_bc and $a'bc$, we have $|s_as_b| = |a'b| \cdot |s_ac|/|a'c| = |a'b| \cdot y(c)/(y(b) - y(c))$, which implies that

$$|a'b| \leq \frac{2A^*}{y(c)^2} \cdot (y(b) - y(c)). \quad (2)$$

Since the upper bound on $|a'b|$ is increasing in $y(c)$ in (1) and decreasing in $y(c)$ in (2), the minimum of the two upper bounds is maximal when they are equal; that is, when $y(c) = y(a)y(b)/(y(a) + y(b))$. It follows that

$$|a'b| \leq \frac{2A^*}{y(a)^2} \cdot (y(a) + y(b)). \quad (3)$$

The condition (3) implies that b is inside a trapezoid of height $y(a)$ with bases of length $4A^*/y(a)$ and $8A^*/y(a)$, which has area $6A^*$. This finishes case 1).

We now consider case 2). Refer to Figure 4b). Let $a'a''$ be the maximum subsegment of s that is visible from a . Since the triangle $aa'a''$ is contained in P we have

$$\text{area}(aa'a'') = \frac{1}{2}|a'a''| \cdot y(a) \leq A^*.$$

If b sees a , then the segment ab intersects the segment $a'a''$. Thus b is contained in a trapezoid of height $y(a)$ with bases of length $|a'a''|$ and $2 \cdot |a'a''|$. Such trapezoid has area

$$\frac{|a'a''| + 2|a'a''|}{2} \cdot y(a) = \frac{3}{2}|a'a''| \cdot y(a) \leq 3A^*.$$

This finishes case 2).

Considering cases 1) and 2) together, for each fixed point $a \in P$ we have

$$\Pr[ab \subset P \text{ and } |y(a)| \geq |y(b)|] \leq 9 \cdot A^*.$$

Since this bound holds for each fixed a , it also holds when a is chosen at random.

Because of symmetry we have

$$\Pr[ab \subset P] = 2 \cdot \Pr[ab \subset P \text{ and } |y(a)| \geq |y(b)|] \leq 18 \cdot A^*,$$

which proves part (i) of the theorem.

Part (ii) follows by a similar consideration using case 2) only. \square

We can use a divide and conquer approach to obtain a bound for arbitrary polygons.

Theorem 9. *Let P be an arbitrary unit-area polygon. Let a and b be two points chosen uniformly at random in P . Then*

$$\Pr[ab \subset P] \leq 12 \cdot A^*(P) \cdot (1 + \log_2(1/A^*(P))).$$

Proof. For this proof, let us set $A^* = A^*(P)$.

For each polygon Q there exists a segment that splits Q into two polygons, each of area at most $\frac{2}{3} \text{area}(Q)$ [9]. We recursively split P using such a segment in each polygon, for $h = \log_{3/2}(1/A^*)$ levels. Thus, at the bottommost level, each polygon has area bounded by A^* .

At each level ℓ of the recursion, where $\ell = 0, \dots, h$, we have 2^ℓ polygons, which we denote by $Q_{\ell,1}, \dots, Q_{\ell,2^\ell}$. In particular, $Q_{0,1} = P$. Since the polygons at each level ℓ are disjoint, we have

$$\sum_{i=1}^{2^\ell} \text{area}(Q_{\ell,i}) = \text{area}(P) = 1.$$

For each polygon $Q_{\ell,i}$, where $\ell < h$, let $e_{\ell,i}$ be the segment used to split $Q_{\ell,i}$. Let $\widehat{Q_{\ell,i}}$ be the portion of $Q_{\ell,i}$ that is weakly visible from $e_{\ell,i}$. At each level $\ell < h$ we have

$$\sum_{i=1}^{2^\ell} \text{area}(\widehat{Q_{\ell,i}}) \leq \sum_{i=1}^{2^\ell} \text{area}(Q_{\ell,i}) = 1.$$

Let $\mathcal{E}_{a,b,\ell,i}$ be the event $ab \subset \widehat{Q_{\ell,i}}$ & $ab \cap e_{\ell,i} \neq \emptyset$. Using the union bound and part (ii) of Theorem 8 we obtain

$$\begin{aligned}
\Pr \left[\bigcup_{\ell=0}^{h-1} \bigcup_{i=1}^{2^\ell} \mathcal{E}_{a,b,\ell,i} \right] &\leq \sum_{\ell=0}^{h-1} \Pr \left[\bigcup_{i=1}^{2^\ell} \mathcal{E}_{a,b,\ell,i} \right] \\
&= \sum_{\ell=0}^{h-1} \sum_{i=1}^{2^\ell} \Pr[\mathcal{E}_{a,b,\ell,i}] \\
&= \sum_{\ell=0}^{h-1} \sum_{i=1}^{2^\ell} \Pr[\mathcal{E}_{a,b,\ell,i} \mid a \in \widehat{Q_{\ell,i}}, b \in \widehat{Q_{\ell,i}}] \cdot \Pr[a \in \widehat{Q_{\ell,i}}, b \in \widehat{Q_{\ell,i}}] \\
&\leq \sum_{\ell=0}^{h-1} \sum_{i=1}^{2^\ell} \left(6 \cdot \frac{A^*}{\text{area}(\widehat{Q_{\ell,i}})} \cdot (\text{area}(\widehat{Q_{\ell,i}}))^2 \right) \\
&= 6 \cdot A^* \sum_{\ell=0}^{h-1} \sum_{i=1}^{2^\ell} \text{area}(\widehat{Q_{\ell,i}}) \\
&\leq 6 \cdot A^* \sum_{\ell=0}^{h-1} 1 \\
&= 6 \cdot A^* \cdot h.
\end{aligned}$$

At the bottommost level h , we can use that $\text{area}(Q_{h,i}) \leq A^*$ for each i to obtain

$$\begin{aligned}
\Pr \left[\bigcup_{i=1}^{2^h} [ab \subset Q_{h,i}] \right] &= \sum_{i=1}^{2^h} \Pr[ab \subset Q_{h,i}] \\
&\leq \sum_{i=1}^{2^h} \Pr[a \in Q_{h,i}, b \in Q_{h,i}] \\
&= \sum_{i=1}^{2^h} (\text{area}(Q_{h,i}))^2 \\
&\leq \sum_{i=1}^{2^h} A^* \cdot (\text{area}(Q_{h,i})) \\
&= A^*.
\end{aligned}$$

We then note that, if a sees b , then the event $\mathcal{E}_{a,b,\ell,i}$ occurs for some $\ell < h$ and $i \leq 2^\ell$, or a and b are in the same polygon $Q_{h,i}$, where $i \leq 2^h$. Thus

$$\begin{aligned}
\Pr[ab \subset P] &\leq \Pr \left[\bigcup_{\ell=0}^{h-1} \bigcup_{i=1}^{2^\ell} \mathcal{E}_{a,b,\ell,i} \right] + \Pr \left[\bigcup_{i=1}^{2^h} [ab \subset Q_{h,i}] \right] \\
&\leq 6 \cdot A^* \cdot h + A^* \\
&= A^* + 6 \cdot A^* \cdot \log_{3/2}(1/A^*) \\
&\leq A^* + 12 \cdot A^* \cdot \log_2(1/A^*).
\end{aligned}$$

□

Algorithm LARGE POTATO**Input:** Unit-area polygon P , $\varepsilon \in (0, 1)$, and $\delta \in (0, 1)$

1. find a value $A(P)$ such that $A(P) \leq A^*(P) \leq C_2 \cdot A(P)$;
2. $r \leftarrow 60/A(P)$;
3. $best \leftarrow \emptyset$;
4. **repeat** $3 \log_2(1/\delta)$ times
 5. $R \leftarrow$ sample r points uniformly at random in P ;
 6. **if** $G(P, R)$ has at most $C_3 \cdot n \log_2 n$ edges **then**
 7. compute $G(P, R)$;
 8. **for** $ab \in E(G(P, R))$ **do**
 9. $R_{ab} \leftarrow$ sample $96 \cdot C_1 \cdot C_2 / (\varepsilon/2)^{3/2}$ points uniformly at random in the parallelogram $\Gamma(a, b, C_2 \cdot A(P))$;
 10. $S_{ab} \leftarrow$ sample $288 \cdot C_2 / \varepsilon$ points uniformly at random in the parallelogram $\Gamma(a, b, C_2 \cdot A(P))$;
 11. $G_{ab} \leftarrow G(P, (R_{ab} \cup S_{ab}) \cap P)$;
 12. **for** $s \in S_{ab}$ **do**
 13. $U \leftarrow \varphi(G_{ab}, s)$;
 14. **if** $\text{area}(U) > \text{area}(best)$ **then** $best \leftarrow U$;
 15. **return** $CH(best)$;

Figure 5: Algorithm. The constant C_1 is from Lemma 1. The constant C_2 is the approximation factor from Hall-Holt et al. [24]; see Section 4.1. The constant C_3 is from Lemma 10.

4 Algorithm

In this section we discuss the eventual algorithm. The input to the algorithm is a polygon P , which without loss of generality we assume that has unit area, a parameter $\varepsilon \in (0, 1)$, and a parameter $\delta \in (0, 1)$. The algorithm, called LARGE POTATO, is summarized in Figure 5. In the first part of the section we explain in detail each step and the notation that is still undefined. In the second part we analyze the algorithm.

4.1 Description

Sampling points. Let $A(P)$ be a constant-factor approximation for $A^*(P)$. Thus, $A(P) \leq A^*(P) \leq C_2 A(P)$ for some constant $C_2 \geq 1$. Hall-Holt et al. [24] provide an algorithm to compute such value $A(P)$ in $O(n \log n)$ time.

Let us define $r = \frac{60}{A(P)}$. Since the largest triangle in any triangulation of P has area at least $1/n$, we have $A^*(P) \geq 1/n$ and thus $r = O(n)$.

Let R be a sample of r points chosen independently at random from the polygon P . The sample R can be constructed in $O(n + r \log n)$ time, as follows. By the linear-time algorithm of Chazelle [11], we compute a triangulation of P , giving triangles T_1, \dots, T_{n-2} . We then compute the prefix sums $S_i = \text{area}(T_1) + \dots + \text{area}(T_i)$ for $i = 1, \dots, n-2$. This is done in $O(n)$ time. To sample a point, we select a random number x in the interval $[0, 1]$, perform a binary search to find the smallest index j such that $x \leq S_j$, and sample a random point inside T_j . A random point inside T_j can be generated using a random point inside a parallelogram that contains two congruent copies of T_j ; such a point can be generated using two random numbers in the interval $[0, 1]$. In total, each point takes $O(\log n)$ time plus the time needed to generate three random numbers in the interval

$[0, 1]$. A similar approach is described in [15].

Size of the visibility graph. Using the expected number of edges in the visibility graph $G(P, R)$ and Markov's inequality leads to the following bound.

Lemma 10. *There exists a constant $C_3 > 0$ such that, with probability at least $5/6$, the graph $G(P, R)$ has at most $C_3 \cdot n \log_2 n$ edges.*

Proof. In this proof we use $G = G(P, R)$. Using linearity of expectation, Theorem 9 and the estimate $A^*(P) \geq 1/n$, we obtain

$$\begin{aligned} \mathbb{E}[|E(G)|] &= \binom{r}{2} \cdot \Pr[\text{two random points are visible in } P] \\ &\leq \frac{1}{2} \left(\frac{60}{A(P)} \right)^2 \cdot 12 \cdot A^*(P) \cdot (1 + \log_2(1/A^*(P))) \\ &\leq 21600 \cdot \frac{(1 + \log_2(1/A^*(P)))}{A^*(P)} \\ &\leq 21600 \cdot n \log_2 n. \end{aligned}$$

Let us take $C_3 = 6 \cdot 21600$. By Markov's inequality we have

$$\Pr[|E(G)| \geq C_3 \cdot n \log_2 n] \leq \frac{\mathbb{E}[|E(G)|]}{C_3 \cdot n \log_2 n} \leq \frac{1}{6}.$$

□

Constructing the visibility graph and checking its size. We will use the following result by Ben-Moshe et al. [8].

Theorem 11 (Ben-Moshe et al. [8]). *Let P be a simple polygon with n vertices and let R be a set of r points inside P . The visibility graph $G(P, R)$ can be constructed in time $O(n + r \log r \log(rn) + k)$, where k is the number of edges in $G(P, R)$.*

In line 6 of the algorithm LARGEPOATATO, we want to check whether $G(P, R)$ has at most $C_3 \cdot n \log_2 n$ edges. For this we use that the algorithm of Theorem 11 is output-sensitive and takes time $T_{[8]}(n, r, k) = O(n + r \log r \log(rn) + k)$. We run the algorithm of Theorem 11 for at most $T_{[8]}(n, r, C_3 \cdot n \log_2 n)$ steps. If the construction of $G(P, R)$ is not finished, we know that $|E(G(P, R))| > C_3 \cdot n \log_2 n$. Otherwise the algorithm outputs whether $|E(G(P, R))| \leq C_3 \cdot n \log_2 n$ or not. Thus, the test in line 6 can be made in time

$$\begin{aligned} T_{[8]}(n, r, C_3 \cdot n \log_2 n) &= O(n + r \log r \log(rn) + C_3 \cdot n \log_2 n) \\ &= O(n + n \log^2 n + n \log n) \\ &= O(n \log^2 n). \end{aligned}$$

The construction in line 7 takes the same time, if it is actually made.

Work for each edge ab . We now discuss the work done in lines 9–14 for each edge ab of $G(P, R)$. The parallelogram $\Gamma(a, b, C_2 \cdot A(P))$ was defined in Section 2.2. Note that $\Gamma(a, b, C_2 \cdot A(P))$ has area

$$12 \cdot C_2 \cdot A(P) \leq 12 \cdot C_2 \cdot A^*(P) = \Theta(A^*(P)).$$

Since Γ_{ab} is a parallelogram, it is straightforward to construct the random samples R_{ab} and S_{ab} . Note that $|R_{ab}| = \Theta(\varepsilon^{-3/2})$ and $|S_{ab}| = \Theta(\varepsilon^{-1})$. We select the subset of $R_{ab} \cup S_{ab}$ contained in the polygon P and construct its visibility graph G_{ab} . We then compute a maximum-area convex clique in G_{ab} restricted to those whose highest vertex c is from S_{ab} . We make this restriction to reduce the number of tries candidate highest points from $\Theta(\varepsilon^{-3/2})$ to $\Theta(\varepsilon^{-1})$. This is equivalent to computing $\varphi(G_{ab}, a)$, which is discussed in Section 2.3. Finally, we compare the feasible solutions U_{ab} that we obtain against the solution stored in the variable *best* and, if appropriate, update *best*.

4.2 Analysis

Lemma 12 (Time bound). *For each ε , where $\frac{1}{n} < \varepsilon < 1$, the algorithm LARGE POTATO can be adapted to use $O((n/\varepsilon^4) \log^2 n \log(1/\delta))$ time.*

Proof. The value A can be computed in time $O(n \log n)$, as discussed before.

We first preprocess the polygon P for segment containment using the algorithm of Chazelle et al. [12]: after $O(n)$ preprocessing time we can answer whether a query segment is contained in P in $O(\log n)$ time. In particular, we can decide in $O(\log n)$ time whether a query point is in P .

We claim that each iteration of the for-loop (lines 9–14) takes $O((1/\varepsilon^4) \log n)$ time. The samples R_{ab} and S_{ab} can be constructed in $O(|R_{ab}| + |S_{ab}|) = O(1/\varepsilon^{3/2})$. We construct $(R_{ab} \cup S_{ab}) \cap P$ by testing each point of $R_{ab} \cup S_{ab}$ for containment in P . The graph G_{ab} is constructed by checking for each pair of points from $(R_{ab} \cup S_{ab}) \cap P$ whether the corresponding segment is contained in P . Thus G_{ab} is constructed in $O((1/\varepsilon^{3/2})^2 \log n) = O((1/\varepsilon^3) \log n)$ time. Each iteration of the lines 13–14 take time $O(|R_{ab}|^2) = O(1/\varepsilon^3)$ time because of Lemma 7. Thus the running time of the for loop in lines 12–14 takes time $O(|S_{ab}| \cdot (1/\varepsilon^3)) = O(1/\varepsilon^4)$. The claim follows.

We next show that each iteration of the repeat-loop (lines 5–14) takes $O((n/\varepsilon^4) \log^2 n)$ time. Since $r = O(n)$, the sample R can be computed in $O(n \log n)$ time, as discussed in Section 4.1. As discussed before, we can make the test in line 6 in $O(n \log^2 n)$ time.

If $G(P, R)$ has more than $C_3 \cdot n \log_2 n$ edges, this finishes the time spent in the iteration. Otherwise, we make $O(n \log n)$ iterations of the for-loop in lines 9–14. Since each iteration of the for-loop takes $O((1/\varepsilon^4) \log n)$ time, as argued earlier in this proof, the bound per iteration of the repeat-loop follows. \square

Lemma 13 (Correctness of one iteration). *In one iteration of the repeat-loop (lines 5–14) of the algorithm LARGE POTATO the algorithm finds a convex polygon of area at least $(1 - \varepsilon)A^*(P)$ with probability at least $1/4$.*

Proof. Let K^* be a polygon contained in P of largest area. Therefore $\text{area}(K^*) = A^*(P)$. Consider one iteration of the repeat-loop. Essentially, there are three things that can go wrong. We define the following events:

$$\begin{aligned} \mathcal{E}_{K^*} : & \text{ for some edge } ab \text{ of } G(P, R), K^* \text{ is contained in } \Gamma(a, b, C_2 \cdot A(P), \\ \mathcal{E}_G : & |E(G(P, R))| \leq C_3 \cdot n \log_2 n, \\ \mathcal{E}_\Gamma : & \text{ for some edge } ab \text{ of } G(P, R), \text{ there is } s \in S_{ab} \text{ such that} \\ & \text{area}(\varphi(G_{ab}, s)) \geq (1 - \varepsilon) \cdot A^*(P). \end{aligned}$$

Since

$$|R| = \frac{60}{A(P)} \geq \frac{60}{\text{area}(K^*)}$$

and $A^*(P) \leq C_2 \cdot A(P)$, Lemma 3 implies that

$$\Pr[\mathcal{E}_{K^*}] \geq \frac{2}{3}.$$

Because of Lemma 10 we have

$$\Pr[\mathcal{E}_G] \geq \frac{5}{6}$$

and therefore

$$\Pr[\mathcal{E}_{K^*} \text{ and } \mathcal{E}_G] \geq \frac{1}{2}. \quad (4)$$

For the rest of the proof, we assume that \mathcal{E}_{K^*} and \mathcal{E}_G hold. Let a_0b_0 be the edge of $G(P, R)$ such that $\Gamma_0 = \Gamma(a_0, b_0, C_2 \cdot A(P))$ contains K^* . The algorithm executes the code in lines 9–14 for $ab = a_0b_0$. Let $K_{\varepsilon/2}^*$ be the portion of K^* above $y = y_{1-\varepsilon/2}(K)$ and let $K_{1-\varepsilon/2}^* = K^* \setminus K_{\varepsilon/2}^*$. It holds that

$$\text{area}(K_{\varepsilon/2}^*) = (\varepsilon/2) \cdot \text{area}(K^*) \quad \text{and} \quad \text{area}(K_{1-\varepsilon/2}^*) = (1 - \varepsilon/2) \cdot \text{area}(K^*).$$

The bound

$$|R_{a_0b_0}| \geq \frac{288 \cdot C_2}{\varepsilon} = 4 \cdot 3 \cdot \frac{12 \cdot C_2 \cdot A^*(P)}{(\varepsilon/2) \cdot A^*(P)} \geq 4 \cdot 3 \cdot \frac{\text{area}(\Gamma_0)}{\text{area}(K_{\varepsilon/2}^*)}$$

and Lemma 2 (with $P = \Gamma_0$ and $K = K_{\varepsilon/2}^*$) imply that

$$\Pr[S_{ab} \cap K_{\varepsilon/2}^* \neq \emptyset \mid \mathcal{E}_{K^*} \text{ and } \mathcal{E}_G] \geq \frac{5}{6}. \quad (5)$$

The bound

$$|R_{a_0b_0}| \geq \frac{96 \cdot C_1 \cdot C_2}{(\varepsilon/2)^{3/2}} = 4 \cdot \frac{C_1}{(\varepsilon/2)^{3/2}} \cdot \frac{12 \cdot C_2 \cdot A^*(P)}{A^*(P)/2} \geq 4 \cdot \frac{C_1}{(\varepsilon/2)^{3/2}} \cdot \frac{\text{area}(\Gamma_0)}{\text{area}(K_{1-\varepsilon/2}^*)}$$

and Lemma 3 (with $P = \Gamma_0$ and $K = K_{1-\varepsilon/2}^*$) imply that

$$\Pr[\text{area}(CH(R_{a_0b_0} \cap K_{1-\varepsilon/2}^*)) \geq (1 - \varepsilon/2) \cdot \text{area}(K_{1-\varepsilon/2}^*) \mid \mathcal{E}_{K^*} \text{ and } \mathcal{E}_G] \geq \frac{2}{3}.$$

Noting that

$$\begin{aligned} \text{area}(CH(R_{a_0b_0} \cap K_{1-\varepsilon/2}^*)) &\geq (1 - \varepsilon/2) \cdot \text{area}(K_{1-\varepsilon/2}^*) \\ &= (1 - \varepsilon/2) \cdot (1 - \varepsilon/2) \cdot A^*(P) \\ &\geq (1 - \varepsilon) \cdot A^*(P), \end{aligned}$$

we have

$$\Pr[\text{area}(CH(R_{a_0b_0} \cap K_{1-\varepsilon/2}^*)) \geq (1 - \varepsilon) \cdot A^*(P) \mid \mathcal{E}_{K^*} \text{ and } \mathcal{E}_G] \geq \frac{2}{3}. \quad (6)$$

Joining (5) and (6) we obtain that, with probability at least $1/2$, it holds

$$\text{area}(CH(R_{a_0b_0} \cap K_{1-\varepsilon/2}^*)) \geq (1 - \varepsilon) \cdot A^*(P) \quad \text{and} \quad S_{ab} \cap K_{\varepsilon/2}^* \neq \emptyset.$$

If s is a point of $S_{ab} \cap K_{\varepsilon/2}^*$, then

$$\begin{aligned} \text{area}(\varphi(G_{a_0b_0}, s)) &\geq \text{area}((K_{1-\varepsilon/2}^* \cap R_{a_0b_0}) \cup \{s\}) \\ &> \text{area}(K_{1-\varepsilon/2}^* \cap R_{a_0b_0}) \\ &\geq (1 - \varepsilon) \cdot A^*(P). \end{aligned}$$

We conclude that

$$\Pr[\mathcal{E}_\Gamma \mid \mathcal{E}_{K^*} \text{ and } \mathcal{E}_G] \geq \frac{1}{2}$$

and using (4) obtain

$$\Pr[\mathcal{E}_{K^*} \text{ and } \mathcal{E}_G \text{ and } \mathcal{E}_\Gamma] = \Pr[\mathcal{E}_\Gamma \mid \mathcal{E}_{K^*} \text{ and } \mathcal{E}_G] \cdot \Pr[\mathcal{E}_{K^*} \text{ and } \mathcal{E}_G] \geq \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

When \mathcal{E}_{K^*} , \mathcal{E}_G and \mathcal{E}_Γ occur, the test in line 6 is satisfied and in one of the iterations of the loop in lines 13–14 we will obtain a $(1 - \varepsilon)$ -approximation to $A^*(P)$. \square

Theorem 14. *Let P be a polygon with n vertices, let ε and δ be parameters with $0 < \varepsilon < 1$ and $0 < \delta < 1$. In time $O((n/\varepsilon^4) \log^2 n \log(1/\delta))$ we can find a convex polygon contained in P that, with probability at least $1 - \delta$, has area at least $(1 - \varepsilon) \cdot A^*(P)$.*

Proof. We consider the output K given by $\text{LARGE POTATO}(P, \varepsilon, \delta)$. Because of Lemma 12, we can assume that the output is computed in time $O((n/\varepsilon^4) \log^2 n \log(1/\delta))$.

The polygon K returned by $\text{LARGE POTATO}(P, \varepsilon, \delta)$ is always a convex polygon contained in P . We have $\text{area}(K) < (1 - \varepsilon) \cdot A^*(P)$ if and only if all iterations of the repeat-loop (lines 5–14) fail to find such a $(1 - \varepsilon)$ -approximation. Since each such iteration fails with probability at most $3/4$ due to Lemma 13, and there are $3 \log_2(1/\delta)$ iterations, we have

$$\Pr[\text{area}(K) < (1 - \varepsilon) \cdot A^*(P)] \leq \left(\frac{3}{4}\right)^{3 \log_2(1/\delta)} < \left(\frac{1}{2}\right)^{\log_2(1/\delta)} = \delta.$$

\square

5 Conclusions

There are several directions for future work. We explicitly mention the following:

- Finding a deterministic $(1 - \varepsilon)$ -approximation using near-linear time.
- Achieving subquadratic time for polygons with an unbounded number of holes.

In the conference version of this paper (to appear in the Proceedings of SoCG 2014), we also mentioned the following two questions that have been answered affirmatively by M. Balko, M. Eliáš, V. Jelínek, P. Valtr and B. Walczak in the meantime; their manuscript is in preparation.

- Does Theorem 8(i) hold for arbitrary simple polygons? We conjecture so, possibly with a larger constant.
- Are similar results achievable in 3-dimensions?

References

- [1] P. K. Agarwal and M. Sharir. Efficient randomized algorithms for some geometric optimization problems. *Discrete & Computational Geometry*, 16(4):317–337, 1996.
- [2] P. K. Agarwal, M. Sharir, and S. Toledo. Applications of parametric searching in geometric optimization. *J. Algorithms*, 17(3):292–318, 1994.

- [3] B. Aronov, M. J. van Kreveld, M. Löffler, and R. I. Silveira. Peeling meshed potatoes. *Algorithmica*, 60(2):349–367, 2011.
- [4] A. Baddeley, I. Bárány, R. Schneider, and W. Weil. *Stochastic Geometry*, volume 1892 of *Lecture notes in mathematics*. Springer, 2007.
- [5] I. Bárány. Random points and lattice points in convex bodies. *Bull. Amer. Math. Soc. (N.S.)*, 45(3):339–365, 2008.
- [6] I. Bárány and D. G. Larman. Convex bodies, economic cap coverings, random polytopes. *Mathematika*, 35:274–291, 1988.
- [7] C. Bautista-Santiago, J. M. Díaz-Báñez, D. Lara, P. Pérez-Lantero, J. Urrutia, and I. Ventura. Computing optimal islands. *Oper. Res. Lett.*, 39(4):246–251, 2011.
- [8] B. Ben-Moshe, O. A. Hall-Holt, M. J. Katz, and J. S. B. Mitchell. Computing the visibility graph of points within a polygon. In J. Snoeyink and J.-D. Boissonnat, editors, *Proc. 20th ACM Symposium on Computational Geometry*, pages 27–35, 2004.
- [9] P. Bose, J. Czyzowicz, E. Kranakis, D. Krizanc, and A. Maheshwari. Polygon cutting: Revisited. In J. Akiyama, M. Kano, and M. Urabe, editors, *Discrete and Computational Geometry*, volume 1763 of *Lecture Notes in Computer Science*, pages 81–92. Springer Berlin Heidelberg, 2000.
- [10] J.-S. Chang and C.-K. Yap. A polynomial solution for the potato-peeling problem. *Discrete & Computational Geometry*, 1:155–182, 1986.
- [11] B. Chazelle. Triangulating a simple polygon in linear time. *Discrete & Computational Geometry*, 6(1):485–524, 1991.
- [12] B. Chazelle, H. Edelsbrunner, M. Grigni, L. Guibas, J. Hershberger, M. Sharir, and J. Snoeyink. Ray shooting in polygons using geodesic triangulations. *Algorithmica*, 12(1):54–68, 1994.
- [13] B. Chazelle, L. J. Guibas, and D. T. Lee. The power of geometric duality. *BIT*, 25(1):76–90, 1985.
- [14] B. Chazelle and M. Sharir. An algorithm for generalized point location and its applications. *J. Symb. Comput.*, 10(3/4):281–310, 1990.
- [15] O. Cheong, A. Efrat, and S. Har-Peled. Finding a guard that sees most and a shop that sells most. *Discrete & Computational Geometry*, 37(4):545–563, 2007.
- [16] K. Daniels, V. Milenkovic, and D. Roth. Finding the largest area axis-parallel rectangle in a polygon. *Computational Geometry*, 7(12):125–148, 1997.
- [17] A. DePano, Y. Ke, and J. O’Rourke. Finding largest inscribed equilateral triangles and squares. In *Proc. 25th Allerton Conf. Commun. Control Comput.*, pages 869–878, 1987.
- [18] A. Dumitrescu, S. Har-Peled, and Cs. D. Tóth. Minimum convex partitions and maximum empty polytopes. In F. V. Fomin and P. Kaski, editors, *SWAT*, volume 7357 of *Lecture Notes in Computer Science*, pages 213–224. Springer, 2012. See <http://arxiv.org/abs/1112.1124> for a longer version.

- [19] H. Edelsbrunner, J. O'Rourke, and R. Seidel. Constructing arrangements of lines and hyperplanes with applications. *SIAM J. Comput.*, 15(2):341–363, 1986.
- [20] B. Efron. The convex hull of a random set of points. *Biometrika*, 52:331–343, 1965.
- [21] P. Fischer. Sequential and parallel algorithms for finding a maximum convex polygon. *Comput. Geom.*, 7(3):187–200, 1997.
- [22] J. E. Goodman. On the largest convex polygon contained in a non-convex n -gon, or how to peel a potato. *Geometriae Dedicata*, 11(1):99–106, 1981.
- [23] H. Groemer. On the mean value of the volume of a random polytope in a convex set. *Archiv der Mathematik*, 25:86–90, 1974.
- [24] O. A. Hall-Holt, M. J. Katz, P. Kumar, J. S. B. Mitchell, and A. Sityon. Finding large sticks and potatoes in polygons. In *Proc. 17th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 474–483, 2006.
- [25] D. T. Lee and Y. T. Ching. The power of geometric duality revisited. *Inform. Process. Lett.*, 21(3):117–122, 1985.
- [26] E. A. Melissaratos and D. L. Souvaine. Shortest paths help solve geometric optimization problems in planar regions. *SIAM J. Comput.*, 21(4):601–638, 1992.
- [27] A. Rényi and R. Sulanke. Über die konvexe Hülle von n zufällig gewählten Punkten. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 2:75–84, 1963.
- [28] A. Rényi and R. Sulanke. Über die konvexe Hülle von n zufällig gewählten Punkten. II. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 3:138–147, 1964.
- [29] G. Rote. The degree of convexity. In S. P. Fekete, editor, *Abstracts 29th European Workshop on Computational Geometry*, pages 69–72, 2013.